

# Classifying discourse in a CSCL platform to evaluate correlations with Teacher Participation and Progress

Eliana Scheihing, Matthieu Vernier, Javiera Born, Julio Guerra, Luis Cárcamo

**Abstract**—In Computer-Supported learning, monitoring and engaging a group of learners is a complex task for teachers, especially when learners are working collaboratively: *Are my students motivated? What kind of progress are they making? Should I intervene? Is my communication and the didactic design adapted to my students?* Our hypothesis is that the analysis of natural language interactions between students, and between students and teachers, provide very valuable information and could be used to produce qualitative indicators to help teachers' decisions. We develop an automatic approach in three steps (1) to explore the discursive functions of messages in a CSCL platform, (2) to classify the messages automatically and (3) to evaluate correlations between discursive attitudes and other variables linked to the learning activity. Results tend to show that some types of discourse are correlated with a notion of Progress on the learning activities and the importance of emotive participation from the Teacher.

**Index Terms**—Collaborative learning tools, Consequences for learning, Discourse, Machine Learning.

## 1 INTRODUCTION

### 1.1 ICT challenges in Education

The emergence of massive open online courses (*Coursera*, *edX*, *OpenClassrooms*), e-learning platforms (*Moodle*, *Khan-Academy*) and Computer-Supported Collaborative Learning platforms (CSCL) (*Metafora*, *Claroline*) offers a new horizon in terms of knowledge building and sharing. At the same time, Blended Learning [1] is a growing pedagogical program in formal education combining traditional classroom sessions and distance learning via web platforms such as the *Kelluwen project* [33]. Several reflections accompany this paradigm shift in Education. We regroup these lines of research under four challenges:

- 1) how to optimize the motivation and satisfaction of learners ? (also called *engagement*)
- 2) how to optimize the skills acquisition and the performance of learners ? (also called *progress*)
- 3) how to predict learning behaviours (engagement / discouragement, progress / mental block, learning style, etc.) ?
- 4) how to optimize the role of the teacher ?

To address these challenges, existing IT research is experimenting and evaluating several ideas:

- 1) gamifying learning experiments [2] [3] [4],
- 2) personalizing learning experiments according to learner profiles [5] [6] [7],
- 3) improving the portability of learning experiments on mobile devices and connected objects to facilitate activities out of classroom [8] [9],
- 4) adding visual feedbacks to increase the learners awareness of their behaviours [10] [11] [12],
- 5) developing interactions and collaborative learning between learners [13], [14], [15].

In the majority of these researches, learners' progress and engagement are evaluated on the basis of structured data available in the platform: test results, time spent on the platform, number of activities realized, number of messages posted or awards received, etc. These indicators give valuable quantitative approximations of progress and engagement levels but are limited to provide a more qualitative understanding of these cognitive phenomena.

### 1.2 The Natural Language: A source of data to explore the cognitive phenomena related to Learning

In the context of CSCL, special attention is given to the development of interactions between learners. In the collaboration theory, it is assumed that knowledge is not only a question of memorising a list of facts, but is constructed dynamically through social relationships, shared activities and dialogues [16]. For this reason, social media (forums, chat rooms, microblogs, video-conferences) are particularly central in CSCL platforms [17] and socio-communicative skills play a key role in knowledge building.

By analysing dialogues between learners in CSCL platforms, previous studies have shown that various linguistic and cognitive phenomena influence or are correlated with progress and engagement of learners:

- 1) mutual listening and dialogical cohesion [18] [19],
- 2) expression of positive/negative emotions [20] [21],
- 3) expression of uncertainty and questioning [22] [23],
- 4) quality of argumentation and vocabulary used [24] [25] [26],
- 5) expression of agreement/disagreement and conflict resolution [27],
- 6) expression of leadership and work organization between learners [28],
- 7) teacher's active participation in discussions [29].

Some of these studies have adopted an automated approach, using Natural Language Processing and Machine Learning techniques [30]. However, significant efforts are still required, on the one hand, to automate textual analysis of complex discursive phenomena in the context of Education (expression of opinion, doubt, belief, disagreement, arguments, etc.), and on the other hand, to evaluate how this new data could be used to produce relevant qualitative indicators and provide assistance in teachers' decision making.

### 1.3 Overview of the research problem addressed in this article

In this article we automate the analysis of the learners' and teachers' discourse in a CSCL platform and then study how discursive phenomena are correlated with other variables such as the quantitative participation of teachers, the level of progress of the learning activities, or the didactic design used. We believe that our work can be used to better support teachers in monitoring the development of learning activities, as well as support instructional designers in developing better learning activities.

Our methodology is divided in three steps. First, we use machine learning techniques to explore latent topics and discover discursive categories present in the messages from learners of teachers. Secondly, we automatically classify the messages according to their main discursive category. Finally, we explore correlations between different variables related to the learning activities and the discursive behaviours observed in the previous step.

The data used in our experiments was taken from the CSCL platform Kelluwen [33]. Since 2011, Kelluwen web platform is used in formal curriculum of public education in southern Chile by about 5000 students in 182 classes of middle school and high school. It includes a microblogging tool called the Logbook (Figure 1) where students and teachers can interact with classmates or other students following the same activity in other schools [34].

The paper is organized as follows. The next section briefly presents useful work for our methodology in the fields of Natural Language Processing (NLP), Machine Learning (ML) and Linguistics. Section 3 describes the Kelluwen platform and dataset on which our experiments are based. In the section 4, 5 and 6, we explain the three steps of our methodology and precise intermediate results for each step. Finally, section 7 summarises our conclusions.

## 2 RELATED WORK

### 2.1 Related work in NLP

In Natural Language Processing (NLP), Latent Dirichlet allocation (LDA) [31] is a topic modeling method often used for identifying thematic from a large volume of unlabeled documents. The general assumption in LDA is that each document is generated from a latent set of probable topics, and each topic is a probability distribution among the entire vocabulary of the collection. LDA analyses co-occurrence of words in the documents to generate a set of topics, each of them as a distribution of probabilities over the vocabulary, by following two optimization goals: minimize



Fig. 1: Overview of the microblogging tool (the "Logbook") in the Kelluwen web platform during a collaborative learning activity in a chilean middle school : *Building a slideshow about the second half of the 20th century*.

the number of probable topics related to each document (describe each document as concisely as possible) and minimize the number of highly probable words within each topic (makes topics as specific as possible). By grouping words that occur together, LDA has the advantage to deal with synonyms and polysemy, which is of great use in our case where we have a considerable amount of miswritten words and use of slang.

Support Vector Machines (SVM) are a set of supervised machine learning techniques that has been widely and successfully used for text classification [32]. SVM were originally designed for binary classification, but it is also able to process multiclass tasks by constructing a set of binary classifiers and using a strategy "one versus the rest" or "one versus one". During the learning phase, from a sample of training examples (for which the category is known), SVM builds a representation of the examples as points in a high-dimensional space. Then, it calculates a mathematical function, the SVM model, separating and optimizing the margin between the points of both classes.

During the test phase, new examples are mapped into that same space and predicted to belong to a category based on which side of the model they fall on. SVM models have been particularly experienced and evaluated in the context of subjectivity analysis and opinion mining to decide if a text is subjective/objective [35] or if it is positive/negative/neutral [36]. In this kind of issues dealing with the discursive level of language, SVM is interesting because it enables to manage a high number of dimensions of different natures.

Finally, approaches combining LDA and SVM have been explored in several fields for classifying the stylometry (linguistic style) [37] or oral transcripts [38] for example. This strategy has the interest of not presupposing any specific categories while taking advantage of SVM. In the particular context of Education, the recent work of Dascalu [30] presents promising results to automate discourse analysis with NLP and ML. His work focus on textual cohesion in dialogues between learners and shows that mutual listening and vocabulary sharing are qualitative indicators of engagement and progress in CSCL.

## 2.2 Related works in Linguistics

In Linguistics, the concept of Discourse implies that the true meaning of a sentence can't be assigned by its only linguistic construction but is largely determined by the communication situation within which it was enounced. In other words, to properly understand a sentence, a first step is to figure out what its main function is in the communication situation and to answer the question, *'With what intention is this message transmitted?'*. The well-known model of the functions of language introduced by Roman Jakobson [39] (Figure 2) defined six main communication functions, according to which an effective act of verbal communication can be described:

- 1) the *emotive* function: the intention of the message is focused on the emotions and attitudes of the speaker towards what he is talking about. Example: *"I have trouble understanding the exercise"*
- 2) the *conative* function: it puts the emphasis on the interlocutor seeking to influence his behaviour by an order or by persuasion. Example: *"Move into groups of two or three"*
- 3) the *phatic* function: the intention is focused on creating a communicative channel, verifying that the channel works properly or simply drawing attention. Example: *Hello! Can you hear me ?*
- 4) the *metalingual* function: it aims to verify that the interlocutors understand each other, speak the same code or agree on the meaning of a particular word. Ex.: *"What do you mean by 'statistics'?"*
- 5) the *referential* function: it aims to give an information about the interlocutors context. Ex.: *"The whale is a mammal"*
- 6) the *poetic* function: it brings to light the message itself, the choice of certain words, their order, their sound, etc. This function aims to express the pleasure and the art of communicating.

Later, other models have been developed to specify subcategories for each of these functions. For example,

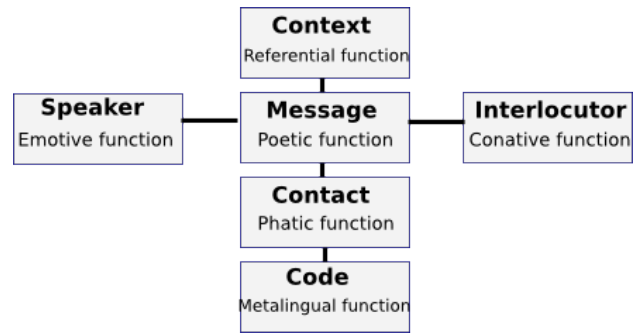


Fig. 2: Jakobson's semiotic theory model describing the 6 main functions of language in verbal communication.

Charaudeau [40] distinguishes several categories in the emotive function such as expression of a "belief", an "agreement/disagreement", a "judgement", etc.

In the context of Education, little scientific research has been published to classify messages according to their discursive functions. Our hypothesis is that this kind of analysis could provide qualitative information: (1) to measure, or even predict, learner's progress and engagement, (2) to improve the teacher awareness on the development of the activity and (3) to evaluate the effects produced by teachers' communication.

## 3 THE KELLUWEN PROJECT AND THE DATASET

The project Kelluwen is an attempt to incorporate the effective use of Blended Learning in formal curriculum of public education in southern Chile. It specifically targets the development of socio-communicative skills of middle and high school students and the use of a computer-supported collaborative tools as didactic resources. It includes a CSCL web platform<sup>1</sup> designed to support the learning activities by providing access to material and tools to facilitate the communication and monitoring of the work (Fig. 1).

### 3.1 The Kelluwen Web platform

In the project, a series of *Didactic Designs* (DD) has been developed to work on three subjects: Language, History and Technology. A DD covers a unit of content as specified in the Chilean public school curriculum, and contains a series of activities which blend traditional classroom learning strategies with the use of social web tools like Wordpress, Youtube, etc. In this way, Kelluwen aims to bring the sharing and communication capabilities of the social web to formal learning activities in schools.

A typical DD contains a sequence of 8 to 12 classroom activities of 1 and 1/2 hours of duration (typical lecture time). In the DD, the students work in teams with the goal of creating an outcome that is finally published in a social Web tool, enabling further peer review processes. To make this happen, activities are organized into three stages:

- 1) a *motivation stage*, where students are introduced to concepts and topics, and they must study the topic and begin to incorporate the knowledge;

1. Available under free license at <http://app.kelluwen.cl>

- 2) a *creation stage*, where using the learned concepts they must generate new content, such as a blog or a video, exercising their communicative skills;
- 3) and an *evaluation stage*, where the generated content is now passed to other students, which may be from the same class group or from a different class group (even from a different school) working in the same DD, who must evaluate and give feedback on the work done. They also have the opportunity to do an auto-evaluation and receive comments from the guiding teacher.

During the activities, students and teachers have access to the Kelluwen platform. Although the main interaction between students and teachers happens in person, students and teachers tend to largely use it in various ways.

### 3.2 The dataset

We collected microblogging posts of 110 class groups working on 31 different Didactic Designs (DD) between 2011 and 2015 (during this period, the Logbook use was optional). Finally, the dataset contains 21.212 messages written in Chilean Spanish. The average length of a message is 7.3 words and 44.8 characters, making it quite short even for a microblog post. Additionally, since the logbook was explicitly intended as an informal channel, the messages contains heavy use of slang, emoticons and miswritten words. The messages are distributed according to the following criteria:

- 20128 messages from students, 1084 from teachers ;
- 9631 from middle school, 11581 from high school ;
- 11341 messages expressed during 51 didactic experiences in Language, 8226 in History (43 experiences), 1655 in Technology (17 experiences).

Students' messages distribution in class groups has a median of 135, mean of 185, a standard deviation of 180, and a maximum of 1104 messages in one of the class groups. Teachers' distribution has a median of 5, mean of 10, and a maximum of 56 messages in one of the groups.

### 3.3 Data preprocessing

In order to facilitate the machine learning processes presented in the sections 4 and 5, we developed a series of natural language preprocessing to normalize the representation of the messages.

- 1) *Tokenization*: a message is segmented into words ;
- 2) *Lemmatization*: the words are annotated with their generic forms (the lemma). For example, the lemma of *actividades* (activities) is *actividad* (activity). For this module, we used the lemmatization model for spanish available in the TreeTagger tool [41],
- 3) *Filtering of stop-words*: a short list of grammatical words (such as *aquel* (that), *cada* (each)) are filtered to simplify the vectorial representation of messages,
- 4) *Basic spell correction*: in the same idea, several slight misspellings are corrected using regular expressions. In particular, it concerns some words containing several vowels running such as

*buuenooooo* (goooooood) which are corrected to *bueno* (good),

- 5) *Emoticons detection*: emoticons (;-), <3, :D, etc.) are detected with regular expressions and are considered like normal words,
- 6) *URL generalization*: all URL links are substituted by the word *http*.

All these preprocessing modules<sup>2</sup> have been developped within the Apache framework UIMA<sup>3</sup> (Unstructured Information Management Architecture).

## 4 EXPLORING TOPICS IN MESSAGES

The first step of our methodology is mainly exploratory. We adopt a LDA non-supervised approach to uncover latent topics in the dataset. The goal is to obtain a first insight on the content of learners' and teachers' messages to help us understand the linguistic attitudes on the Kelluwen CSCL platform.

### 4.1 Topic extraction with LDA

LDA requires to postulate the number of clusters (or topics) as a parameter. After applying the preprocessing described in the section 3.3, we ran the topic modeling for 3, 4, 5 and 6 clusters using *R* and the package *lda* with the Gibbs sampling algorithm<sup>4</sup>. We repeat each experiment 3 times and compute log-likelihood.

Results of maximum log-likelihood run for the experiment with 3 topics are shown in the table 1 where the Top-20 most probable words within each topic are shown. A review of the probabilities within each topic reveals that the Top-20 words cover 34% (for Topics 2 and 3) and 39% (for Topic 1) of the topic, respectively. On initial examination, we observe that Topic 3 contains a lot of words related to the content of didactic designs (*world, literary, water, etc.*). When repeating the process for 4, 5 and 6 clusters, Topic 1 and Topic 2 remain the same, and the rest of the topics split the words accordingly to the content of different didactic designs. In other words, LDA tends to rediscover the various learning activities when the number of clusters is set from 4 to 6. Because we aim to analyze the topics regardless of specific contents, we decided to consider 3 as a sufficient number of clusters to analyze and interpret these first results.

### 4.2 Discussion on the exploratory results

In table 1, we observe that Topic 1 includes opinion words (*good, interesting, entertaining*) and words that refer to the collaborative learning activity (*work, classmate, activity*). Therefore, we believe that Topic 1 relates to the messages when a learner or a teacher express a subjective attitude about the learning experience or about the work of others. In the Jakobson theory, this kind of messages typically refers to the Emotive function of language (Fig. 2). Topic 2

2. The dataset and preprocessing modules are availables at the following address under free license: <https://github.com/matthieuviernier/>.

3. <http://uima.apache.org/>

4. parameters used: burn-in=100 000 and iterations=500.000.

Topic 1 (Emotive)	Topic 2 (Phatic)	Topic 3 (Referential)
trabajo ( <i>work</i> )	XD (EMOTICON)	mundo ( <i>world</i> )
bueno ( <i>good</i> )	tú ( <i>you</i> )	literario ( <i>literary</i> )
compañero ( <i>classmate</i> )	ola ( <i>hello</i> )	texto ( <i>text</i> )
actividad ( <i>activity</i> )	hola ( <i>hello</i> )	http (URL)
bien ( <i>nice</i> )	po ( <i>hi</i> )	cotidiano ( <i>daily</i> )
parecer ( <i>to seem</i> )	wena ( <i>cool</i> )	real ( <i>real</i> )
gustar ( <i>to like</i> )	holi ( <i>hi</i> )	ser ( <i>to be</i> )
aprender ( <i>to learn</i> )	:\$ (EMOTICON)	NUMERIC VALUE
interesante ( <i>interesting</i> )	:p (EMOTICON)	hecho ( <i>done</i> )
entretenido ( <i>entertaining</i> )	jajaja ( <i>lol</i> )	vida ( <i>life</i> )
mucho ( <i>very much</i> )	callar ( <i>to be quiet</i> )	porque ( <i>because</i> )
terminar ( <i>to finish</i> )	cabro ( <i>boy</i> )	onirico ( <i>oneiric</i> )
experiencia ( <i>experience</i> )	si ( <i>yes</i> )	historia ( <i>history</i> )
kelluwen ( <i>kelluwen</i> )	fome ( <i>boring</i> )	mitico ( <i>mythical</i> )
esperar ( <i>to hope/to wait</i> )	papá ( <i>dad</i> )	ejemplo ( <i>example</i> )
trabajar ( <i>to work</i> )	:) (EMOTICON)	fantástico ( <i>fantastic</i> )
nuestro ( <i>our</i> )	:c (EMOTICON)	agua ( <i>water</i> )
grupo ( <i>group</i> )	:b (EMOTICON)	pertenecer ( <i>belong to</i> )
subir ( <i>to climb/to raise</i> )	:l (EMOTICON)	realidad ( <i>reality</i> )
estar ( <i>to be</i> )	oli ( <i>hi</i> )	del ( <i>of the</i> )

TABLE 1

Distribution of the main words (and an approximate translation) in the three topics according to LDA with their associated function of the language (Emotive, Phatic and Referential).

contains greeting words (*hello*, *hi*), interlocutor markers (*you*, *boy*) and emoticons. This topic corresponds to the many informal messages written on the platform. Those messages are not directly linked to the didactic activity, they are used to establish and maintain a communication channel (called Phatic function in the Jakobson theory), mainly between students. We also note the significant use of emoticons in this group of words. Emoticons are usually only considered as opinion markers, but we argue that emoticons could also have an important phatic function, in particular among young learners. As we said before, Topic 3 regroups words directly related with the content of didactic designs. It also contains a lot of URLs (*http*) indicating that this topic concerns the work itself. This topic is used to share information about a particular context, it is what Jakobson called the Referential function of Language.

Finally, we note that LDA has not identified topics that refers to the the Poetic, Metalingual and Conative functions of Language. Two reasons may explain this: (1) There is no message for these functions in our context. It is probably the case of the *Poetic* function. (2) The lexical items are the same as those of an other function. It could be the case of the *Conative* function and the *Metalingual* function, which can be confused respectively with the *Phatic* function and the *Referential* function. A more specific study would be necessary to distinguish them.

## 5 CLASSIFYING MESSAGES

In the first step, we explored learners' and teachers' messages and identified they might be grouped in 3 discursive categories: Emotive, Phatic and Referential. The second step consists of evaluating the feasibility of automating the classification of messages. From a reference corpus, we developed and evaluated two statistical models, based on LDA and SVM, to classify messages of the Kelluwen platform.

### 5.1 Reference corpus & Manual annotation

In order to build a reference corpus, 500 messages from the initial dataset have been randomly selected. Then, 3 human judges determined the main discursive function of each message among the 3 functions identified in the previous step (Emotive, Phatic or Referential). We used the Fleiss' kappa [42] to measure the reliability of agreement between judges. For this task, the Fleiss' Kappa is 0.70 which is a good agreement according to the scale proposed by Landis et al. [43]. In other words, even if in a few cases this task can be complex for humans (when messages have a really poor syntax or contain only an emoticon for example), most of the time humans are able to understand with what intention a message was transmitted.

Finally, for each message of the reference corpus, the category is the one chosen by at least two judges. There is no cases where all judges disagreed with the two others. The reference corpus contains 240 messages with a Phatic function, 137 messages with an Emotive function and 123 messages with a Referential function.

### 5.2 Classification methods

#### 5.2.1 SVM

The first method is based on a SVM supervised learning approach. For each message, we apply the preprocessing as described in the section 3.3 to build a vector representation. In addition to the class of the message, the vector attributes are the generic form of words existing in the entire corpus (without the stop-words) which in this case are 1,543 words. The attribute value is binary and indicates if the word occurs or not in the message. For example, the vector representation of the example (1) contains three attributes with the value "1" : no, entender (*to understand*) and ":-(" . Other attributes (the other 1,540 words) have the value "0".

(1) No entiendo :- ( *I don't understand* :- ( (Emotive)

For our experiment, we used the Sequential minimal optimization (SMO) algorithm available in the Weka library<sup>5</sup>. SMO is an algorithm for training SVM, it has the interest to split large SVM learning problems into a series of smaller tasks that can be solved analytically and faster. We trained a SVM model and evaluate it on the reference corpus by using a 10-folds cross-validation technique.

#### 5.2.2 LDA

The second method used to automatically classify the messages is based on the LDA approach presented in the section 4. With LDA, each word has a likelihood of belonging to a given topic. Therefore, this classifier simply calculates the likelihood of a message to be in class  $C_i$  as the product of the likelihood of each word of the message to be in class  $C_i$ . The class with the best likelihood is the one kept by this classifier. Because LDA is not directly adapted for supervised classification, this classifier is mainly used as a baseline to evaluate the SVM classifier.

Class	LDA			SVM		
	P	R	F1	P	R	F1
Phatic	84.9%	59.0%	69.6%	87.6%	97.9%	92.5%
Emotive	74.6%	73.0%	73.8%	100.0%	100.0%	100.0%
Referential	47.2%	76.4%	58.4%	94.7%	73.2%	82.6%
Total	72.8%	67.1%	69.8%	92.8%	92.4%	92.6%

TABLE 2

Precision, Recall and F1-Score nmeasures of learners messages classification into three classes with 2 classifiers: LDA and SVM.

### 5.3 Classification results & Discussion

To determine the effectiveness of both classifiers, we use 3 classical measures in Text Classification: (1) the **Precision** (P) measures the system ability to avoid irrelevant solutions, the **Recall** (R) measures the system ability to provide all the relevant solutions and the **F1-Score** (F1) is the harmonic mean between Precision and Recall.

$$P(class_i) = \frac{tp}{tp+fp} \quad R(class_i) = \frac{tp}{tp+fn}$$

$$F1(class_i) = \frac{2PR}{P+R}$$

Where :

- *tp* (true positive): classifier says "class<sub>i</sub>" and judge says "class<sub>i</sub>"
- *fp* (false positive): classifier says "class<sub>i</sub>" and judge says "NOT class<sub>i</sub>"
- *fn* (false negative): classifier says "NOT class<sub>i</sub>" and judge says "class<sub>i</sub>"

A general overview of the results (Table 2) tends to show that discursive classification of short messages on CSCL platform can be achieved automatically even if texts are noisy. In both methods, we observe that the Emotive class obtain the best F1-Score and is perfectly analyzed by SVM. This result may reflect that young learners do not use a wide range of vocabulary or grammatical constructions when expressing their emotions and that the lexical items are quite specifics to the Emotive class. In both methods, the Referential messages represented the lowest F1-Score. This result is not surprising because this category regroups a wider diversity of lexical items related to several didactic design contents. The SVM approach tend to forget relevant Referential messages (the Recall is the lowest for this class) and consider that they are Phatic messages (Precision of the Phatic class is the lowest). One way to improve this aspect could be to collect automatically a set of words related to the didactic design and integrate this knowledge before the classification process.

## 6 EVALUATING CORRELATIONS

In the previous step, we show that a first level of discursive analysis can be achieved automatically on short messages of CSCL platforms with fairly results. In this final step, we investigate how these different types of discourse are correlated or not to other measurable variables.

5. <http://www.cs.waikato.ac.nz/ml/weka/>

	R_S	P_S	E_S	R_T	P_T	E_T	P	S	M	V
Referential_S	-									
Phatic_S	-0.71	-								
Emotive_S	0.10	-0.77	-							
Referential_T	0.11	0.04	-0.16	-						
Phatic_T	-0.19	0.09	0.04	-0.41	-					
Emotive_T	0.06	-0.11	0.09	-0.15	-0.20	-				
Progress	0.12	-0.19	0.15	0.06	0.10	0.02	-			
Students	0.00	0.09	-0.14	0.06	-0.03	0.15	0.23	-		
Mean	-0.10	0.13	-0.10	0.10	0.03	-0.10	-0.39	-0.10	-	
Var	0.04	-0.02	-0.01	-0.05	0.05	-0.02	-0.06	-0.06	0.64	-
Teacher Part.	-0.04	0.03	-0.01	0.20	0.15	-0.05	-0.03	-0.26	0.21	0.08

TABLE 3

Correlation matrix between the discursive attitudes of Students and Teachers and other variables linked to the learning experiences.

### 6.1 Classification of all the messages

With the SVM model that we trained and evaluated in the previous step, we classified the 20,128 messages of students from the dataset. According to this automatic classification, the dataset contains 9,984 Phatic messages (median of 61 per classroom group and standard deviation of 122), 3,341 Referential messages (median: 18, standard deviation: 33) and 7,704 Emotive messages (median: 41, standard deviation: 94). We also classified the 1,084 messages from teachers (44 Phatics, 199 Referentials and 841 Emotives).

For each classroom group, we compute the proportion of each of these categories. These variables are called *Phatic*, *Referential* and *Emotive*, with a suffix *\_S* or *\_T* to indicate if it is the proportion in Students' or Teachers' messages.

### 6.2 Description of the other variables considered

For each group, we also consider the following variables:

- **Participation of teachers:** This variable is computed as the percentage of teacher messages over all messages posted in a class group and is named *Teacher Participation*.
- **Number of Students:** This represents the number of students that belong to each class group. It is not necessarily the same number of students who participate in posting on the microblog. This factor is named *Students*.
- **Progress of the experience:** some class groups never finished the Didactic Design. This variable is computed as the percentage of completed activities in each class group and is called *Progress*.
- **Regularity of the activities:** This is measured through obtaining the differences in days between activities, and determining the mean and variance for each class group, so that a small mean and variance implies regularity. The two factors are called *Mean* and *Variance*.

### 6.3 Correlation results & Discussion

In statistics, the interpretation of a correlation coefficient  $x$  ( $x \subseteq [-1, +1]$ ) is a complex task. It is generally considered that there is a strong correlation when  $|x|$  is close to 0.7, a weak correlation when  $|x|$  is close 0.3 and no correlation when  $|x|$  is close to 0. In our case, as we are working with linguistic phenomena, it can be interesting to observe results with less stringent criteria while remaining critical on the interpretation.



### 6.3.1 Correlation matrix

In the correlation matrix (Tab. 3), we observe that there are several strong correlations between the variables Mean/Variance ( $x=0.64$ ), Phatic\_S/Referential\_S ( $x=-0.71$ ) and Emotive\_S/Phatic\_S ( $x=-0.77$ ). These variables are correlated by definition, but we can still observe how the Phatic variable is negatively related to the two other discursive categories.

A more interesting correlation is the one observed between Progress and students' attitudes: Phatic\_S ( $x=-0.19$ ), Referential\_S ( $x=0.12$ ) and Emotive\_S ( $x=0.15$ ). We may think that chatting (excessive Phatic communication) has a bad influence on the success of the learning experience. On the contrary, Emotive or Referential attitudes from students seem to be slightly related to Progress. We may assume that Referential messages tend to be related to a notion of engagement in the learning experience and that Phatic messages, when they are too preponderant, reveal that learners do not focus their attention on the substance of the learning activity. This analysis could provide an interesting way to quickly detect if the teacher must intervene to refocus the activity.

Furthermore, we observe that quantitative participation of the teacher (Teacher Part.) does not seem to be related to Learners' attitudes : Referential\_S ( $x=-0.04$ ), Emotive\_S ( $x=-0.01$ ) and Phatic\_S ( $x=0.03$ ). Nevertheless, if we look more qualitatively at the teachers' attitudes, we can see that Phatic messages from Teachers tend to be negatively related to Referential messages from Students ( $x=-0.19$ ). On the contrary, Emotive messages from Teachers tend to be negatively related to Phatic messages from Students ( $x=-0.11$ ). In other words, some qualitative participation of the teacher may have a relationship with the learners' behaviours. In particular, young learners seem to be sensitive to Emotive feedback from the teacher.

### 6.3.2 Principal Components Analysis

For a graphical analysis, we generated a Principal Component Analysis (PCA) over the data. PCA is a transformation technique for finding patterns in multidimensional data. The Figure 3 shows a two dimensional representation where correlations can be observed as the arrows point to the same (positive correlations), opposite (negative correlations) or orthogonal (no correlations) directions. Each arrow is one of the considered variable, the numbers positioned in the figure correspond to the didactic design ID of each experiences. For example, it shows that the didactic design 21 (Language activity using Youtube) is often related to Progress, Emotive messages from teachers and Referential messages from students. In this sense, the discursive analysis of messages could also be used to help teachers or designers in choosing and developing better learning activities.

Additionally, this figure confirms a quite strong negative correlation between Progress and the average number of days between activities (variable Mean) ( $x=-0.39$ ). We believe that this lack of regularity reflects the teacher's need to be accompanied by indicators to help him manage the pace of the Computer-Supported learning activity over time. A good way to do that could be to monitor the number of

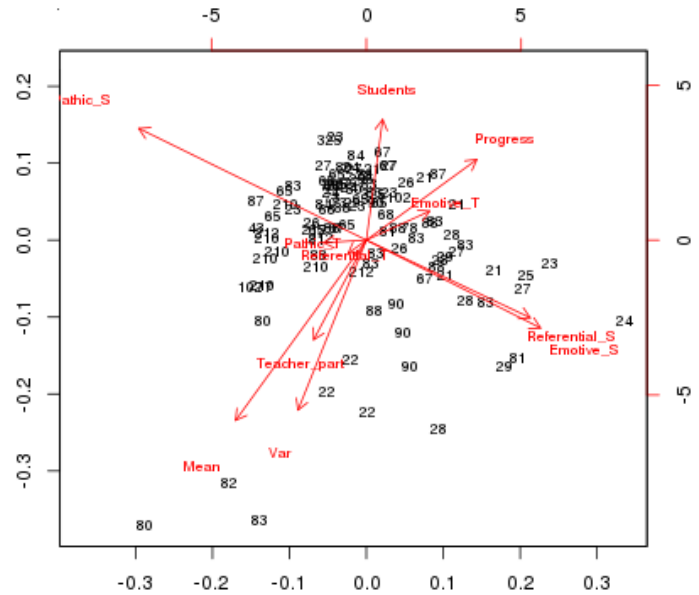


Fig. 3: PCA analysis over the 11 variables described. The main plane represents 47,41% of the data variance.

Phatic/Emotive/Referential messages over time to tell the teacher when it is time to finish an activity, start a new one or give emotive feedback. This figure also tends to confirm the correlation between Progress and the proportion of Emotive communication from the teacher. In other words, encouraging emotive feedback from the teacher would be a key aspect to improve learners' engagement and progress in a Computer-Supported learning environment.

## 7 CONCLUSIONS

In this article, we show that the analysis of Natural Language interactions between students and teachers provide valuable information that could help teachers' decisions in the context of Computer-Supported Learning. We demonstrated the feasibility to automate the analysis of these interactions even if texts are short and noisy. With a combination of non-supervised and supervised machine learning techniques, we identified that messages of Learners and Teachers can be classified automatically according to three intentions of communication described by Jakobson: create or maintain a communication channel (Phatic), share information about the learning context (Referential) and express an emotion or a subjective attitude (Emotive).

The results obtained from a correlation calculation bring to light that the types of discourse used by Learners tend to have a relationship with the Teacher qualitative Participation and the Progress of the learning activity. In particular, Emotive communication from the teacher tends to be an interesting aspect to consider to improve learners' engagement in Computer-Supported learning. In this way, the volume of Phatic/Emotive/Referential messages could be monitored automatically over time to provide a complementary indicator of engagement/disengagement. It can also be used to help the teacher evaluate Learning Designs and his own communication.

## REFERENCES

- [1] B. Gzer, H. Caner, *The Past, Present and Future of Blended Learning: An in Depth Analysis of Literature*, In Proceedings of the 5th World Conference on Educational Sciences, Social and Behavioral Sciences, Vol. 116, pp 45964603, 2014.
- [2] A. Domnguez, J. Saenz De Navarrete, L. De Marcos, L. Fernandez-Sanz, C. Pags, J.-J. Martnez-Herriz, *Gamifying learning experiences: Practical implications and outcomes*, Computers & Education, vol.63, pp.380-392, 2013.
- [3] M.-B. Ibanez, A. Di-Serio, C. Delgado-Kloos, *Gamification for Engaging Computer Science Students in Learning Activities: A Case Study*, IEEE Transactions on Learning Technologies, vol.7-3, pp. 291-301,2014.
- [4] D. Hooshyar, R. Binti Ahmad, M. Yousefi, M. Fathi, S.-J. Horng, H. Lim, *Applying an online game-based formative assessment in a flowchart-based intelligent tutoring system for improving problem-solving skills*, Computers & Education, vol. 94, pp 1836, 2016.
- [5] J. Yang, Z. Xing Huang, Y. Xiang Gao, H. Tao Liu, *Dynamic Learning Style Prediction Method Based on a Pattern Recognition Technique*, IEEE Transactions on Learning Technologies, vol.7-2, pp. 165-177, 2014.
- [6] J. Feldman, A. Monteserin, A. Amandi, *Detecting students' perception style by using games*, Computers & Education, vol. 71, pp 1422, 2014.
- [7] E. Rahimi, J. van den Berg, W. Veen, *A learning model for enhancing the student's control in educational process using Web 2.0 personal learning environments*, British Journal of Educational Technology, vol. 46-4, pp 780792, 2015.
- [8] K. Chin, K. Lee; Y. Chen, *Impact on Student Motivation by Using a QR-Based U-Learning Material Production System to Create Authentic Learning Experiences*, in IEEE Transactions on Learning Technologies, vol.8-4, pp.367-382, 2015.
- [9] C.-K. Looi, D. Sun, W. Xie, *Exploring Students Progression in an Inquiry Science Curriculum Enabled by Mobile Learning*, IEEE Transactions on Learning Technologies, vol.8-3, pp. 43-54, 2015.
- [10] T. Auvinen, L. Hakulinen, L. Malmi, *Increasing Students Awareness of Their Behavior in Online Learning Environments with Visualizations and Achievement Badges*, IEEE Transactions on Learning Technologies, vol.8-3, pp. 261-273, 2015.
- [11] C. Demmans Epp, S. Bull, *Uncertainty Representation in Visualizations of Learning Analytics for Learners: Current Approaches and Opportunities*, IEEE Transactions on Learning Technologies, vol.8-3, pp. 242-260, 2015.
- [12] P. Brusilovsky, Sibel Somyürek, J. Guerra, R. Hosseini, and V. Zadorozhny, *The Value of Social: Comparing Open Student Modeling and Open Social Student Modeling*. In User Modeling, Adaptation and Personalization, pp. 44-55. Springer Intern. Publishing, 2015.
- [13] T. Engelmann, R. Kolodziej, F. W. Hesse, *Preventing undesirable effects of mutual trust and the development of skepticism in virtual groups by applying the knowledge and information awareness approach*, International Journal of CSCL, vol. 9-2, pp 211-235, 2014.
- [14] S. Goggins, W. Xing, *Building models explaining student participation behavior in asynchronous online discussion*, Computers & Education, vol. 94, pp 241251, 2016.
- [15] E. Vanderhoven, A. Raes, H. Montrieux, T. Rotsaert, T. Schellens, *What if pupils can assess their peers anonymously? A quasi-experimental study*, Computers & Education, vol. 81, pp 123132, 2015.
- [16] G. Stahl, *Building collaborative knowing: Elements of a social theory of CSCL*. In J.-W. Stribos, P. Kirschner & R. Martens (Eds.), *What we know about CSCL: And implementing it in higher education*, pp. 53-86, 2004.
- [17] J.J.H.M. Janssen, G. Erkens, G. Kanselaar, *Visualization of agreement and discussion processes during computer-supported collaborative learning*, Computers in Human Behavior, vol. 23-3, pp. 1105-1125, 2007.
- [18] A. F. Wise, S. N. Hausknecht, Y. Zhao, *Attending to others posts in asynchronous discussions: Learners online listening and its relationship to speaking*, International Journal of CSCL, vol 9-2, pp 185-209, 2014.
- [19] M. Dascalu, S. Trausan-Matu, D. S. McNamara, P. Dessus, *ReaderBench: Automated evaluation of collaboration based on cohesion and dialogism*, International Journal of CSCL, vol. 10-4, pp 395-423, 2015.
- [20] Y.-c. Yeh, G.-J. Lai, C. Fu Lind, C.-W. Line, H.-C. Sunf, *How stress influences creativity in game-based situations: Analysis of stress hormones, negative emotions, and working memory*, Computers & Education, vol. 81, pp 143153, 2015.
- [21] X. Chen, M. Vorvoreanu, K. P. C. Madhavan, *Mining Social Media Data for Understanding Students Learning Experiences*, IEEE Transactions on Learning Technologies, vol.7-3, pp. 246-259, 2014.
- [22] Michelle E. Jordan , An-Chih Janne Cheng, Diane Schallert, Kwangok Song, SoonAh Lee, Yangjoo Park, *I guess my question is: What is the co-occurrence of uncertainty and learning in computer-mediated discourse?*, Int'l Journal of CSCL, vol. 9-4, pp 451-475, 2014.
- [23] Y. Zhang, Y. Li, P. Cao, R. Huang, *Understanding Learners' Interaction Patterns through Analyzing Context of Questions in Discussion Transcripts*, IEEE International Conference on Advanced Learning Technologies, pp. 548-552, 2011.
- [24] E. Alagoz, *Social argumentation in online synchronous communication*, Int'l Journal of CSCL, vol. 8-4, pp 399-426, 2013.
- [25] M. Dascalu, T. Rebedea, S. Trausan-Matu, *A Deep Insight in Chat Analysis: Collaboration, Evolution and Evaluation, Summarization and Search*, Artificial Intelligence: Methodology, Systems, and Applications, vol. 6304 of the series Lecture Notes in Computer Science, pp 191-200, 2010.
- [26] G. P. Jain, V. P. Gurupur, J. L. Schroeder, E. D. Faulkenberry, *Artificial Intelligence-Based Student Learning Evaluation: A Concept Map-Based Approach for Analyzing a Student's Understanding of a Topic*, IEEE Transactions on Learning Technologies, vol.7-3, pp. 267-279, 2014.
- [27] A. Weinberger , M. Marttunen, L. Laurinen, K. Stegmann, *Inducing socio-cognitive conflict in Finnish and German groups of online learners by CSCL script*, Int'l Journal of CSCL, vol. 8-3, pp 333-349, 2013.
- [28] E. M. Mercier , S. E. Higgins, L. da Costa, *Different leaders: Emergent organizational and intellectual leadership in childrens collaborative learning groups*, Int'l Journal of CSCL, vol. 9-4, pp 397-432, 2014.
- [29] C. G. Brinton, M. Chiang, S. Jain, H. Lam, Z. Liu, F. M. Fai Wong, *Learning about Social Learning in MOOCs: From Statistical Analysis to Generative Model*, IEEE Transactions on Learning Technologies, vol.7-4, pp. 346-359, 2014.
- [30] M. Dascalu, *Analyzing Discourse and Text Complexity for Learning and Collaborating A Cognitive Approach Based on Natural Language Processing*, Studies in Computational Intelligence, Springer, 2014.
- [31] D. Blei, A. Ng, M. Jordan, *Latent Dirichlet allocation*, Journal of Machine Learning Research, pp 9931022, 2003.
- [32] T. Joachims, *Learning to Classify Text Using Support Vector Machines – Methods, Theory, and Algorithms*, Kluwer/Springer, 2002.
- [33] E. Scheihing, J. Guerra, L. Cárcamo et al., *La experiencia Kelluwen: Tres años de desarrollo y puesta en práctica de una propuesta de innovación didáctica con uso de TIC*. Estudios pedagógicos, vol. 39, pp. 121-141, 2013.
- [34] K. Inalef, J. Guerra, E. Scheihing, *Development and Validation of a Virtual Worklog as a Collaboration Tool for the Kelluwen Learning Community*, Trust, Security and Privacy in Computing and Communications (TrustCom), IEEE 11th International Conference on IEEE, 2012.
- [35] T.A Wilson, *Fine-grained Subjectivity and Sentiment Analysis: Recognizing the intensity, polarity, and attitudes of private states*, Doctoral Dissertation, University of Pittsburgh, 2008.
- [36] M. Rushdi Saleh, M.T. Martn-Valdivia, A. Montejo-Rez, L.A. Urea-Lpez, *Experiments with SVM to classify opinions in different domains*, Journal of Expert Systems with Applications, Vol. 38, pp. 14799-14804, 2011.
- [37] R. Arun, R. Saradha, V. Suresh, M. Narasimha Murty, C. E. Veni Madhavan, *Stopwords and Stylometry: A Latent Dirichlet Allocation Approach*, In NIPS Workshop on Applications for Topic Models : Text and Beyond, 2009.
- [38] M. Morchid, R. Dufour, G. Linarès, *A Lda-Based Topic Classification Approach from Highly Imperfect Automatic Transcriptions*, In: LREC 2014, pp 1309-1314, 2014.
- [39] R. Jakobson, *Linguistics and Poetics*, in T. Sebeok, ed., *Style in Language*, Cambridge, MA: M.I.T. Press, pp. 350-377, 1960.
- [40] P. Charaudeau, *Grammaire du sens et de l'expression*, Paris, Hachette ducation, 927 p., 1992.
- [41] H. Schmid, *Improvements in Part-of-Speech Tagging with an Application to German*. Proceedings of the ACL SIGDAT-Workshop. Dublin, Ireland, 1995.
- [42] J. L. Fleiss, *Measuring nominal scale agreement among many raters*, in Psychological Bulletin, vol. 76, no. 5, pp. 378382, 1971.
- [43] J. R. Landis, G. G. Koch, *The measurement of observer agreement for categorical data*, in Biometrics, vol. 33, pp. 159174, 1977.



a